

# Ekstaoppgaver i STK1110 høsten 2007

## Oppgave E1

Anta  $X \sim \text{bin}(n, p)$ , dvs. binomisk fordelt basert på  $n$  forsøk med suksessannsynlighet  $p$ . Anta at  $n$  er så stor at fordelingen til andelen  $\hat{p} = X/n$  kan tilnærmes med en normalfordeling, dvs. fordelingen til

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \quad (1)$$

er tilnærmet  $N(0, 1)$ . En kan vise at også fordelingen til

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \quad (2)$$

er tilnærmet  $N(0, 1)$ .

- (a) Bruk (2) til å vise at et tilnærmet  $100(1 - \alpha)\%$  konfidensintervall for  $p$  har formen

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \quad (3)$$

der  $z_{1-\alpha/2}$  er  $1-\alpha/2$  kvantilen (fraktilen) til standard normalfordelingen.

- (b) En intervju-undersøkelse av 400 personer viser at 320 har lett tilgang til internett. Forklar når det er rimelig å betrakte dette som et binomisk forsøk. Beregn et 90% konfidensintervall for andelen  $p$  som har lett adgang til internett.
- (c) Hvor stor må  $n$  være for at lengden av konfidensintervallet (3) høyst skal være fem prosentenheter?
- (d) Vi kan få et konfidensintervall for  $p$  som er bedre enn (3) ved å ta utgangspunkt i tilnærmelsen (1) i stedet for tilnærmelsen (2). Bestem et slikt konfidensintervall. (*Vink:* Sannsynligheten er tilnærmet  $1 - \alpha$  for at (1) skal ligge i intervallet  $[-z_{1-\alpha/2}, z_{1+\alpha/2}]$ . Det gir to andregrads ulikheter. Løs disse ulikhetene m.h.p.  $p$ .)

## Oppgave E2

- (a) Lag et lite program i MATLAB der du simulerer realisasjoner av  $X_1, X_2, \dots, X_{10}$ , der  $X_i$ -ene er uavhengige og  $N(5, 1)$ -fordelte. Bruke disse realisasjonene til å lage et 90% konfidensintervall for forventningen. (Når du lager konfidensintervallet, tenker du deg at både forventningen og variansen i normalfordelingen er ukjente.)
- (b) Gjenta punkt a 50 ganger. Finn andelen av de 50 intervallene som inneholder 5 (dvs. den sanne forventningen).
- (c) Forklar resultatet du fant. Antall intervaller som inneholder 5, er en tilfeldig variabel. Hva er fordelingen for denne?

*Vink til punkt b:* Kommandoen `y = normrnd(5, 1, [10, 50])` lager en  $10 \times 50$  matrise med realisasjoner av uavhengige  $N(5, 1)$ -fordelte variable. Hver kolonne i denne matrisa svarer altså til realisasjoner av 10 uavhengige  $N(5, 1)$ -fordelt variable.